Учреждение образования

«БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНОЛОГИЧЕСКИЙ УНИВЕРСИТЕТ»

Расчёт параметров и информационных характеристик дискретных ИС

Студент: Валдайцев А. Д.

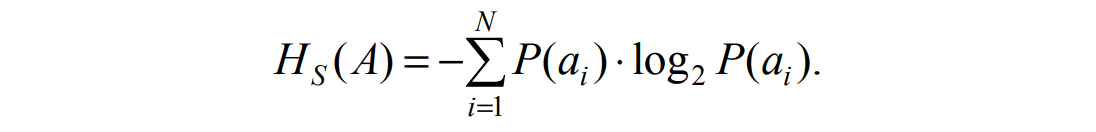
ФИТ 3 курс 5 группа

Преподаватель: Савельева М. Г.

Минск 2023

# Энтропия алфавита

Информационной характеристикой алфавита (источника сообщений на основе этого алфавита) является энтропия. Энтропию алфавита *А = {ai}* по К. Шеннону рассчитывают по следующей формуле:

(1.1)

С физической точки зрения энтропия алфавита показывает, какое количество информации приходится в среднем на один символ алфавита.

Для вычисления энтропии в начале необходимо подсчитать количество появлений каждого символа в исходной строке. Для этого используется функция, представленная на рисунке 1.1.

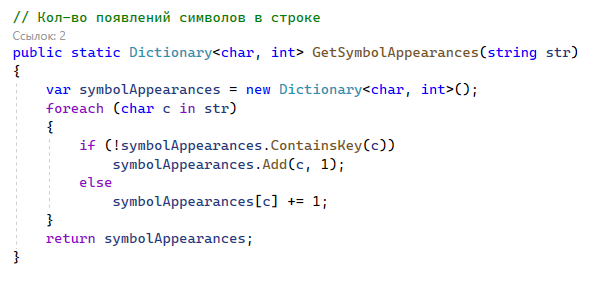


Рисунок 1.1 – Функция подсчёта кол-ва появлений символов

Далее, для вычисления энтропии, необходимо перейти от количества появлений символов к вероятности их появления. Для этого необходимо кол-во появлений каждого символа разделить на кол-во символов в строке. Далее эту вероятность *P(ai)* можно подставлять в формулу (1.1). Функция, вычисляющая энтропию Шеннона, представлена на рисунке 1.2.

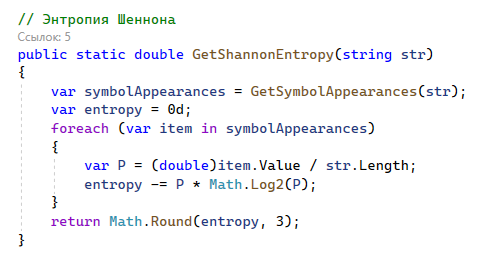


Рисунок 1.2 – Функция вычисления энтропии Шеннона

# Кириллица

В качестве кириллического алфавита был выбран киргизский. Электронный текстовый документ на основе алфавита содержится в файле kyrgyz.txt, представленном на рисунке 1.3.

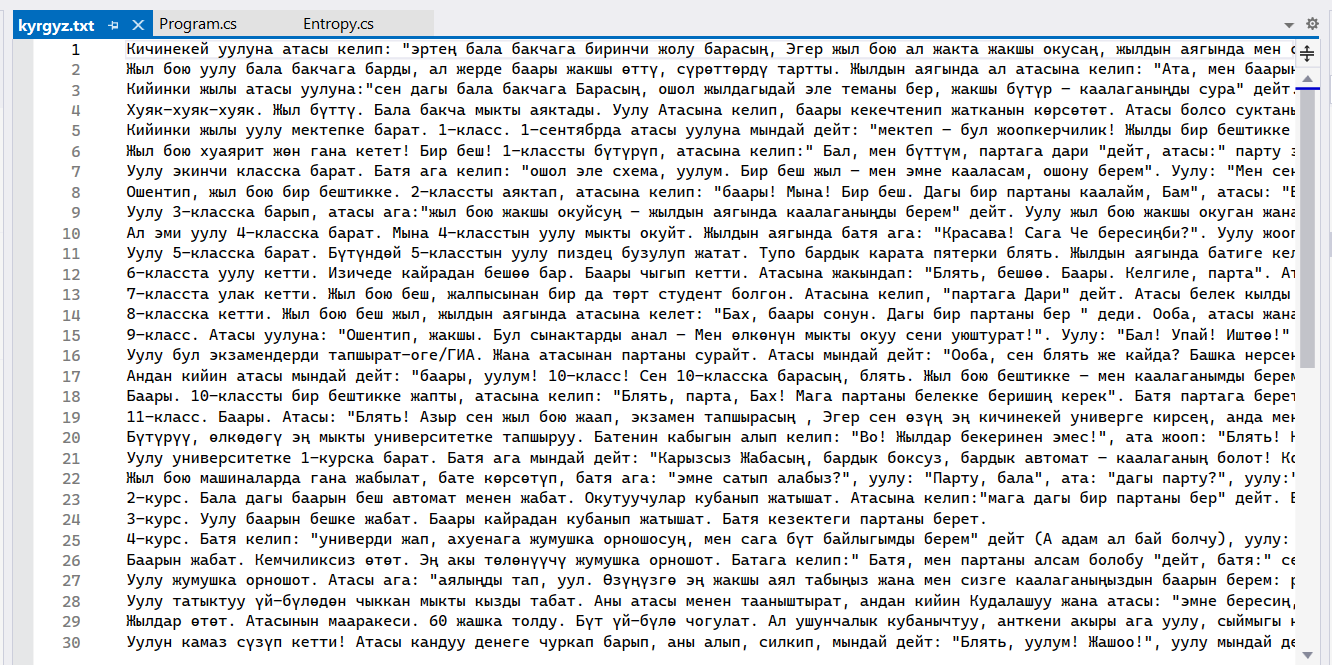


Рисунок 1.3 – Текст на киргизском алфавите

# Латиница

В качестве алфавита на латинице был выбран литовский алфавит. Электронный документ, в котором содержится текст на исходном литовском алфавите для подсчета энтропии алфавита, содержится в файле lithuanian.txt, представленном на рисунке 1.4.

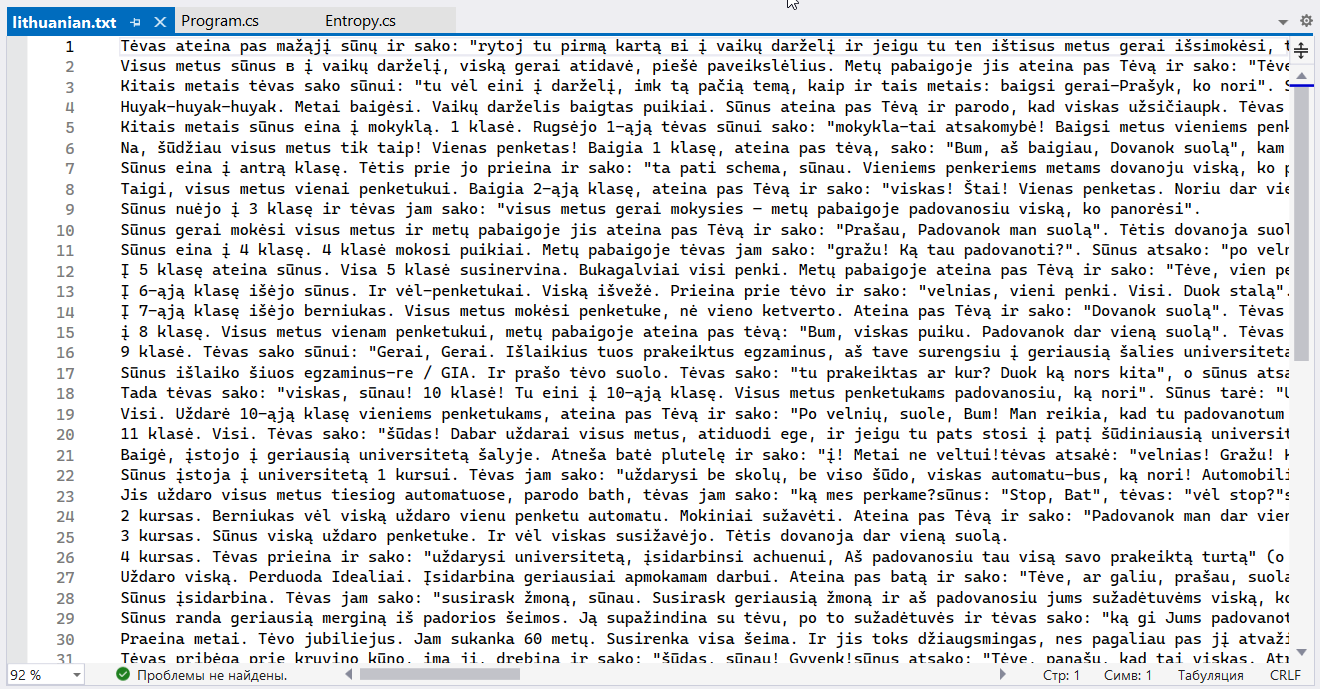


Рисунок 1.4 – Текст на литовском алфавите

# Вычисление энтропии алфавита

Графики, отражающие частоты появления символов в текста на киргизском и литовском алфавитах, представлены на рисунке 1.5.

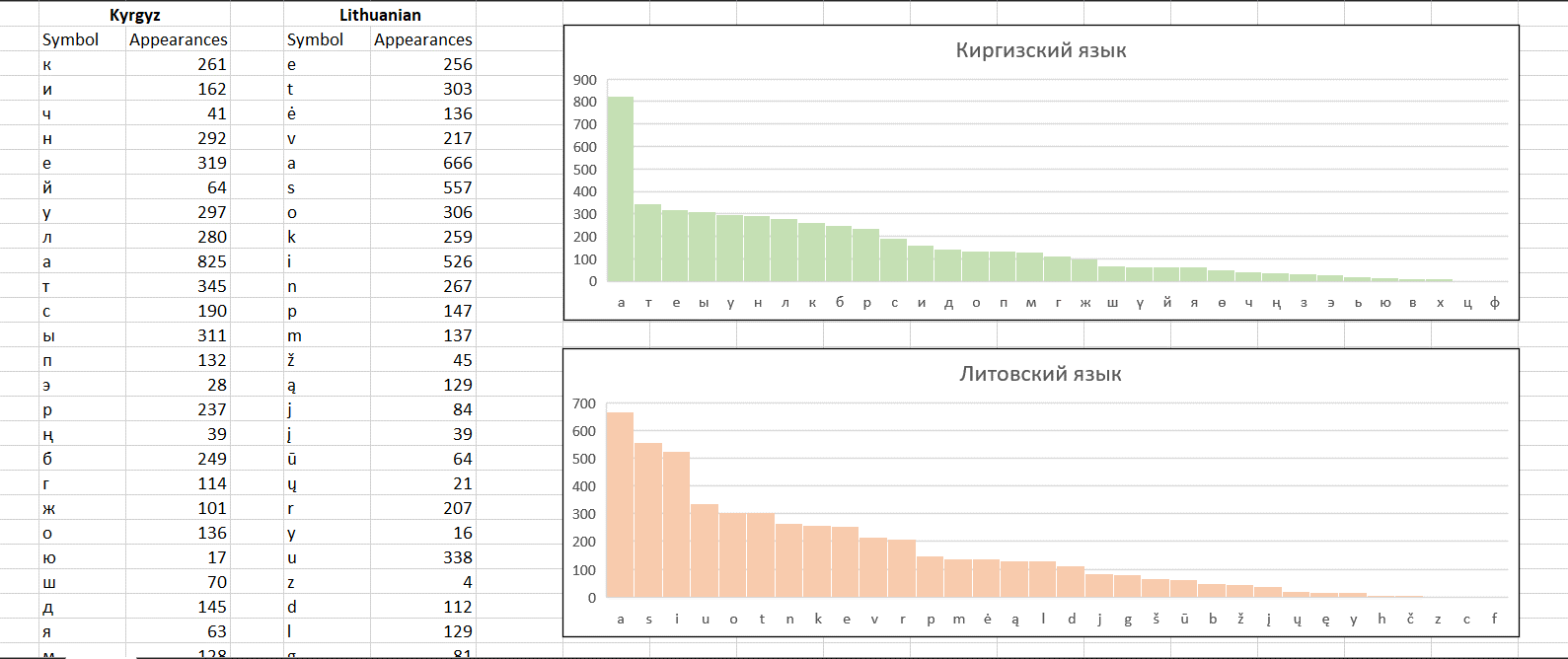


Рисунок 1.5 – Частоты появления символов в алфавитах

Из этих данных можно получить вероятности появления каждого символа *P(ai).* При подстановке их в формулу (1.1) получаем следующие значения энтропии для киргизского, литовского и бинарного алфавитов, продемонстрированные на рисунке 1.6:

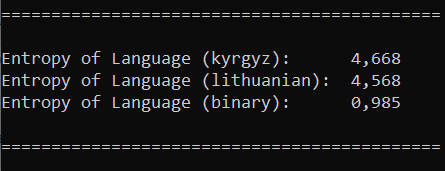


Рисунок 1.6 – Вывод функции подсчёта энтропии

Энтропия киргизского и литовского алфавитов равны соответственно 4,668 бит и 4,568 бит. Энтропия бинарного алфавита, которая всегда примерно равна 1 биту, в данном случае равняется 0,985 бит.

# Количество информации

При известной энтропии алфавита количество информации вычисляется по следующей формуле:

(1.2)

То есть, необходимо умножить полученную энтропию на количество символов в сообщении. Программный код вычисления количества информации сообщения представлен на рисунке 2.1.

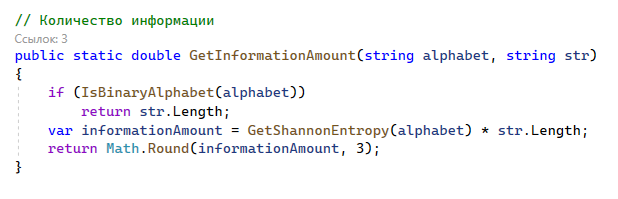


Рисунок 2.1 – Функция количества информации

Если алфавит является бинарным, то за количество информации принимается длина строки, так как энтропия равняется 1. В противном случае, энтропия выбранного алфавита умножается на количество символов в сообщении, количество информации в котором мы желаем найти.

Вывод функции представлен на рисунке 2.2.

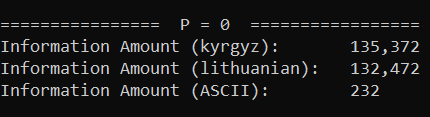


Рисунок 2.2 – Вывод количества информации

Как видно из данного рисунка, количество информации в сообщении, состоящем из ФИО в кодировке ASCII, равняется 232 битам, то есть равняется количеству символов в исходном сообщении, записанном в бинарном виде.

# Эффективная энтропия

Если сообщение может передаться ошибочно, то для вычисления количества информации используется формула эффективной энтропии:

(1.3)

где *H(Y | X)* – условная энтропия:

(1.4)

Программный код функции для вычисления эффективной энтропии представлен на рисунке 3.1.

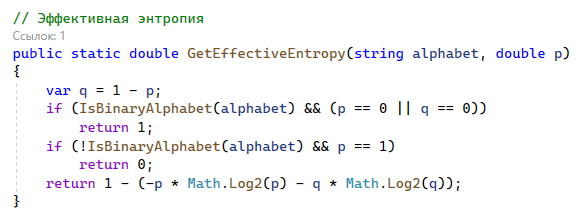


Рисунок 3.1 – Функция эффективной энтропии

В бинарном алфавите эффективная энтропия и, соответственно, количество информации, будет равняться длине строки в случае, если вероятность ошибки равняется 0 или 1. В небинарных же алфавитах, при вероятности ошибки, равной 1, эффективная энтропия и количество информации будет равняться 0.

В таком случае, функция вычисления количества информации станет выглядеть следующим образом:

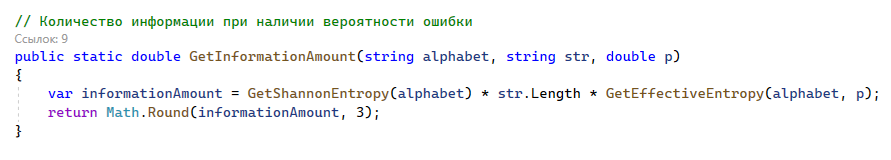


Рисунок 3.2 – Функция кол-ва информации при наличии ошибок

Вывод функции вычисления количества информации при вероятности ошибки, отличной от нулевой, представлен на рисунке 3.3.

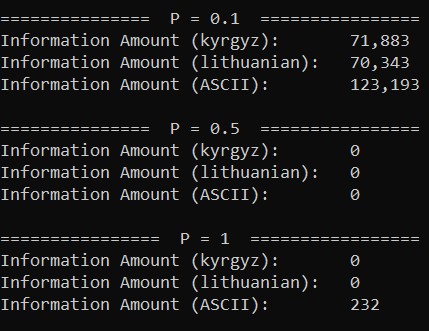


Рисунок 3.3 – Вывод кол-ва информации при *p>0*

**Вывод:** При вероятности ошибки *p=1* количество информации в бинарном алфавите равняется количеству информации при *p=0*. Это связано с тем, что все биты исходного сообщения инвертируются, то есть заменяются на обратные, поэтому из такого сообщения можно получить исходную информацию. В небинарном же алфавите при *p=1* определить исходное сообщение не представляется возможным ввиду того, что мощность алфавита больше 2, и каждый символ заменяется на некоторый неизвестный символ из алфавита, поэтому количество информации в таком случае равняется 0.